Characterizing place: an empirical comparison between user-generated content and freelisting data

Wartmann, Flurina M ; Derungs, Curdin ; Purves, Ross S

Abstract: Methods to gather information from the public about place range from ethnographic approaches such as free listing to automatic extraction from user-generated content. We compared aspects of place (location, locale and sense of place) contained in free lists elicited from participants recruited on site with tags from georeferenced Flickr images. Using manual annotation we assigned content as toponyms (location), landscape elements (locale) and feelings (sense of place). Flickr tags contained more toponyms than free-listing data, but almost no information relating to feelings. Landscape elements were prominent in both data sets, with those captured by free lists and Flickr being cognitively more salient than those only captured by free listing, suggesting they represent basic levels. In Flickr, landscape elements consisted of basic levels in different languages (e.g. mountain, Berg), while free lists contained landscape elements both at the basic and subordinate level (e.g. Arvenwälder, arolla pine forests). We conclude that both methods yielded information about locale, with Flickr contributing basic levels and free lists also more detailed information, but that Flickr provided little information about sense of place compared to in situ free-listing elicitation with participants.

DOI: https://doi.org/10.21433/B3111bw2c801

Characterizing place: an empirical comparison between user-generated content and free-listing data

F. M. Wartmann[1], C. Derungs[1,2], R.S. Purves[1,2]

[1]Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
Email: {flurina.wartmann; curdin.derungs; ross.purves}@geo.uzh.ch

[2]URPP Language and Space, University of Zurich, Freiestrasse 6, CH-8032 Zurich

Abstract

Methods to gather information from the public about place range from ethnographic approaches such as free listing to automatic extraction from user-generated content. We compared aspects of place (*location*, *locale* and *sense of place*) contained in free lists elicited from participants recruited on site with tags from georeferenced Flickr images. Using manual annotation we assigned content as toponyms (*location*), landscape elements (*locale*) and feelings (*sense of place*). Flickr tags contained more toponyms than free-listing data, but almost no information relating to feelings. Landscape elements were prominent in both data sets, with those captured by free lists and Flickr being cognitively more salient than those only captured by free listing, suggesting they represent basic levels. In Flickr, landscape elements consisted of basic levels in different languages (e.g. *mountain, Berg*), while free lists contained landscape elements both at the basic and subordinate level (e.g. *Arvenwälder*, arolla pine forests). We conclude that both methods yielded information about *locale*, with Flickr contributing basic levels and free lists also more detailed information, but that Flickr provided little information about *sense of place* compared to *in situ* free-listing elicitation with participants.

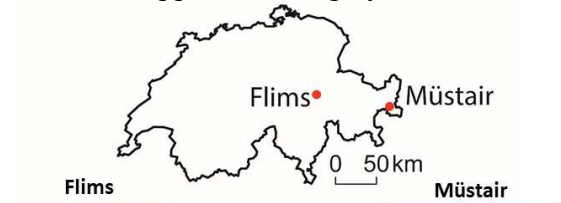
1. Introduction

The importance of taking into account meanings people assign to places in management and planning is increasingly being recognized (Jones 2007; Prieur *et al.* 2006). However, collecting such information from the public is non-trivial, and different approaches are used, ranging from ethnographic work to extraction from user-generated content. Given the diversity of approaches, it is crucial to compare different approaches and understand what types of information are captured by different methods such that research on place can effectively capture relevant information. In this paper, we compare free-listing elicitation with automated extraction of place descriptions from user-generated content. In free-listing experiments, participants are asked to freely name examples of, for instance, categories they associate with the landscape they currently find themselves in (Bieling *et al.* 2014; Wartmann *et al.* 2015). Another method increasingly used is to automatically extract information from user-generated content, often in the form of image tags (e.g. Jenkins *et al.* 2016; Purves *et al.* 2011; Rattenbury and Naaman 2009). We compared these two methods based on the information about place they contained.

2. Methods

We chose two sites in the Swiss Alps (Figure 1) where outdoor free listings on landscape categories had been conducted with 59 participants recruited on site (Wartmann *et al.* 2015). Terms were elicited with a Swiss German question that literally translates to: *'what is there for you in a*

landscape?', resulting in 332 unique terms. From a set of ~1 million georeferenced images from the Flickr platform we automatically extracted image tags using a 10km radius around our study sites. After filtering for bulk-uploads and removing tags only applied by a single user, we retained a total of 633 tags, each of which was applied as a tag by between 2 to 241 users.



Flims

Müstair

Flickr tags	Free-listing data
Laax	*Berge (mountains)*
Flims	*Bäume (trees)*
Graubünden	*Felsen (rocks)*

Flickr tags	Free-listing data
Graubünden	*Berge (mountains)*
Ofenpass	*Tal (valley)*
Müstair	*Wald (forest)*

Figure 1. Study sites and example data (Flickr tags with highest TF-IDF values and free-listing data with highest cognitive saliency index)

We compared the two data sets using three aspects of place as defined by Agnew (1987): *location* (a specific location identified through coordinates, or a name), *locale* (the actual elements that characterize a place, such as *streets*, *buildings*, *rivers*, or *forests*) and *sense of place* (the meanings and feelings people attach to a place). As a language-independent means of comparing the data sets, we manually annotated content types. We devised written annotation guidelines, using the content type 'toponym' for *location*, (e.g. *Flims, Buffalora*), 'landscape element' for *locale*, which included both natural (e.g. *lake, forest, mountain)* and anthropogenic elements (e.g. *road, restaurant*), 'feeling' for *sense of place* (e.g. *peace, quietness*), and 'other' for other content that mostly related to photography, qualities, and person or event names (e.g. *350d, blue, burtonopen*). Three researchers assigned content types using the guidelines and discussed unclear cases until all content was annotated. We calculated the absolute number and percentage of tags/free-listing elements per content type. For the type 'landscape element' we assessed correspondences between Flickr and free lists by considering terms that were identical, translations of each other (e.g. *river-Fluss*), or singular/plural forms (e.g. *river-Flüsse*). For Flickr tags we compared a spatial term frequency–inverse document frequency (TF-IDF, c.f. Rattenbury and Naaman 2009) and for free-listing data cognitive saliency values (Sutrop 2001) between corresponding and non-corresponding terms using Mann-Whitney-U tests to assess statistical significance at a significance level of 0.05.

3. Results and Interpretation

In the following, we briefly present our results with respect to the classification introduced above for Flickr tags and free-listing elements (Table 1). Toponyms are more prevalent in Flickr than our free listings. By contrast, landscape elements are in absolute terms numerous in both data sets, though relatively less prominent in Flickr. Almost no tags relate to feelings in Flickr tags extracted for the study sites, while our free-listing data contain 21 such terms, including *Geborgenheit* (emotional security), *mit sich eins sein* (to be at one with oneself), and *Erdverbundenheit* (connection to the Earth). Both data sets contain relatively high proportions of landscape elements.

Content classified as 'other' dominates Flickr, and includes a wide range of tags which related to, for example, qualities (e.g. *snow, cloudy*), activities (e.g. *hiking, skiing*), and camera metadata.

Table 1. Content types for Flickr tags and free-listing data for the study sites

	Flickr		Free lists	
Flims	(absolute)	(%)	(absolute)	(%)
toponyms	99	19.64%	7	2.41%
landscape elements	60	11.91%	134	46.05%
feelings	1	0.20%	11	3.78%
other	151	68.25%	2	47.77%
Val Müstair				
toponyms	67	31.46%	12	7.55%
landscape elements	28	13.15%	67	42.14%
feelings	0	0.00%	10	6.29%
other	57	55.40%	5	44.03%

As landscape elements are common in both datasets, we explore the degree of overlap between this content, and compare the nature of overlapping and non-overlapping tags/free-listing data. Figure 2 illustrates the properties of Flickr tags and free-listing elements for Val Müstair.

Figure 2. Corresponding and non-corresponding landscape elements in tags and free lists

A number of observations can be made. Firstly, Flickr is dominated by English, while our free-listing (since it was carried out in variant of German) only contains German terms. Secondly, only 8 Flickr tags are not replicated in the free-listing data, while 54 landscape elements were unique to free-listing data. Thirdly, we found correspondences between 19 Flickr tags with 13 free-listing landscape elements. The reason for the mismatch in these counts is illustrated in Figure 2, with for example matches between synonyms (e.g. *creek/stream* in Flickr with *bach* in free lists), as well as between singular/plural forms and different languages (e.g. *mountain/mountains/montagna/ montagne* with *berg/berge*). Since we hypothesized that landscape elements which were more salient in our lists were more likely to have correspondences with Flickr tags, we tested for significant differences in TF-IDF and cognitive salience values between corresponding and not corresponding landscape elements. For Flickr tags (TF-IDF) we found no significant relationship,

while for free-listing landscape elements cognitive salience values were significantly higher for corresponding landscape element for both Flims and Val Müstair.

4. Discussion

Our results confirm previous findings that Flickr tags are a good source for toponyms, that is, in Agnew's terms, for information about *location* (Sigurbjörnsson and van Zwol 2008), but we found very few tags relating to feelings or *sense of place*. Although the free-listing experiment was not aimed at specifically eliciting *sense of place*, participants provided such information. Both data sets contained landscape terms, but due to their specific elicitation, free lists contained more terms despite few participants. Landscape terms in free lists shared with Flickr were significantly more cognitively salient than non-shared terms, suggesting shared terms may be candidates for highly salient basic-level landscape categories (Tversky and Hemenway 1983) such as *river*, or *mountain*, while elements only included in free-listing data often appear to be typical subordinate categories (e.g. *rough pasture* or *mountain torrent*). The shared Flickr terms did not have significantly higher TFIDF-values than non-shared terms, indicating that, as potential basic-level categories, these terms are represented throughout the Flickr corpus (Rorissa 2008), and not particularly overrepresented in the study areas. Flickr was a reliable source of basic-level categories for *locale*, but contained little additional information about landscape elements not present in free-listing data. For future work we thus suggest triangulating methods, where *location* and basic-level categories for *locale* are extracted from user-generated content, with ethnographic methods providing more detailed information about *locale* and, using particular elicitation questions, about *sense of place*.

Acknowledgements

This work was supported by the cogito foundation (grant no. 15-129-R) and the Swiss National Science Foundation (SNF) through the project PlaceGen (grant no. 149823).

References

Agnew JA, 1987, *Place and politics. The geographical mediation of state and society*. Allen & Unwin, Boston, USA.

Bieling C, Plieninger T, Pirker H and Vogl CR, 2014, Linkages between landscapes and human well-being: An empirical exploration with short interviews. *Ecological Economics*, 105:19–30.

Jenkins A, Croitoru A, Crooks AT, Stefanidis A, 2016, Crowdsourcing a collective sense of place. *PLoS ONE*, 11(4).

Jones M, 2007. The European landscape convention and the question of public participation. *Landscape Research*, 32(5):613–633.

Prieur M, *et al.*, 2006, *Landscape and sustainable development-challenges of the European Landscape Convention.* Council of Europe Publishing, Strasbourg, France.

Purves RS, Edwardes A and Wood J, 2011, Describing place through user generated content. *First Monday*, 16(9).

Rattenbury T and Naaman M, 2009, Methods for extracting place semantics from Flickr tags. *ACM Transactions on the Web*, 3(1):1–30.

Rorissa, A, 2008, User-generated descriptions of individual images versus labels of groups of images: A comparison using basic level theory. *Information Processing & Management*, 44(5): 1741–1753.

Sigurbjörnsson B and van Zwol R, 2008, Flickr tag recommendation based on collective knowledge. In *Proceedings of WWW'08*. ACM Press, New York, USA, 327–336.

Sutrop U, 2001, List Task and a Cognitive Salience Index. *Field Methods*, 13(3):263–276.

Tversky B and Hemenway K, 1983, Categories of environmental scenes. *Cognitive psychology*, 15(1):121–149.

Wartmann FM, Egorova E, Derungs C, Purves RS, and Mark DM, 2015, More than a list: what outdoor free listings of landscape categories reveal about commonsense geographic concepts and memory search strategies. In Fabrikant SI, Raubal M, Bertolotto M, Davies C, Freundschuh S, and Bell S (eds.), *Spatial Information Theory. Lecture Notes in Computer Science,* Vol. 9368. Springer International, Cham, Switzerland, 224–243.